

# Recent findings and technological advances in phosphoproteomics for cells and tissues

*Expert Rev. Proteomics* 12(5), 469–487 (2015)

Louise von Stechow,  
Chiara Francavilla and  
Jesper V Olsen\*

*Proteomics Program, The Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark*  
\*Author for correspondence:  
jesper.olsen@cpr.ku.dk

Site-specific phosphorylation is a fast and reversible covalent post-translational modification that is tightly regulated in cells. The cellular machinery of enzymes that write, erase and read these modifications (kinases, phosphatases and phospho-binding proteins) is frequently deregulated in different diseases, including cancer. Large-scale studies of phosphoproteins – termed phosphoproteomics – strongly rely on the use of high-performance mass spectrometric instrumentation. This powerful technology has been applied to study a great number of phosphorylation-based phenotypes. Nevertheless, many technical and biological challenges have to be overcome to identify biologically relevant phosphorylation sites in cells and tissues. This review describes different technological strategies to identify and quantify phosphorylation sites with high accuracy, without significant loss of analysis speed and reproducibility in tissues and cells. Moreover, computational tools for analysis, integration and biological interpretation of phosphorylation events are discussed.

**KEYWORDS:** bioinformatics analyses • mass spectrometry • phospho-enrichment strategies • phosphoproteomics • quantitative phosphoproteomics

## Current status & future challenges in the phosphoproteomics field

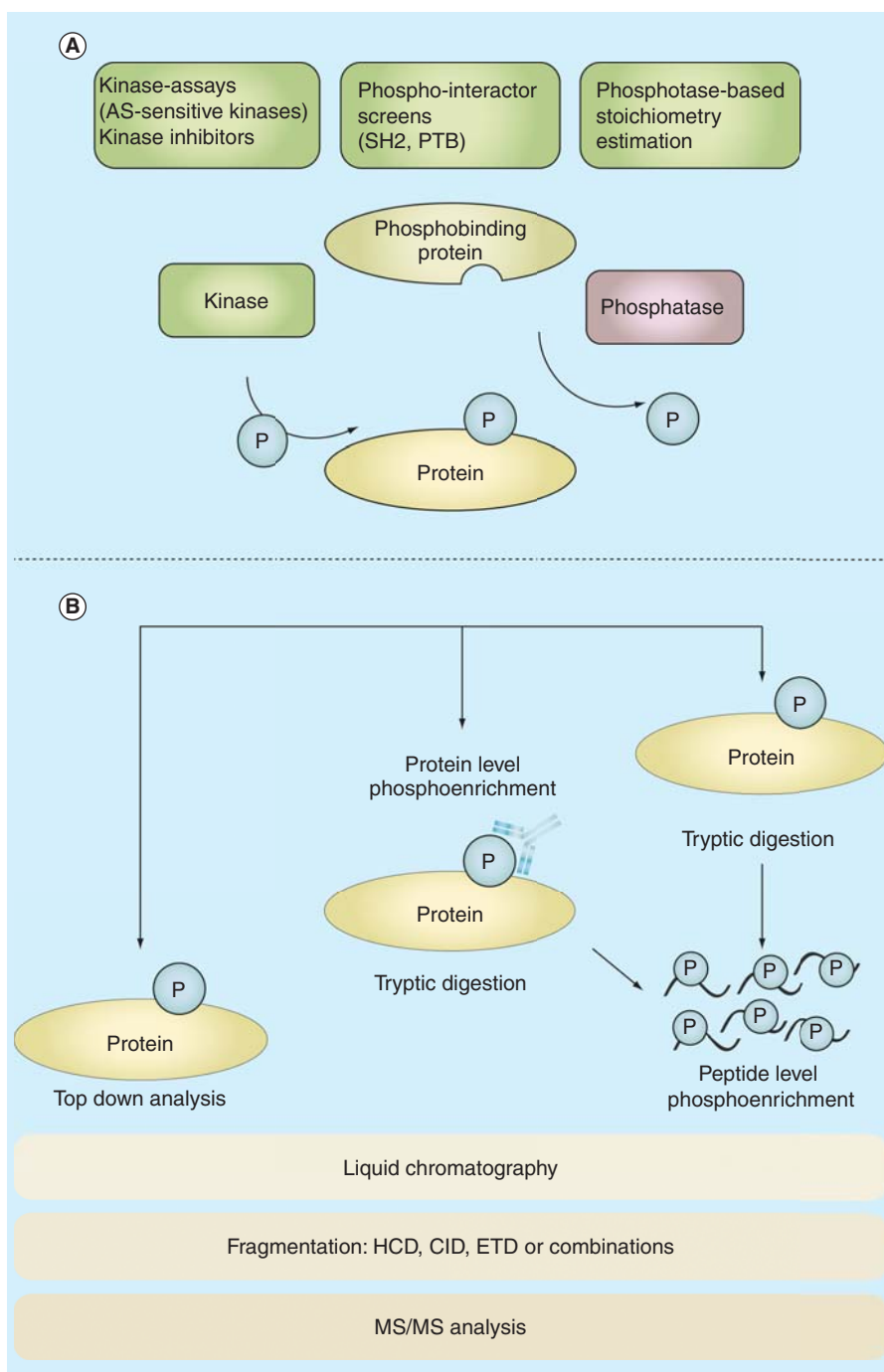
Site-specific protein phosphorylation is a fast and reversible posttranslational modification (PTM). PTMs constitute the major currency of cellular signaling events and are crucial regulators of virtually all cellular processes. Deregulation of phosphorylation-mediated cellular signaling pathways is associated with different diseases including cancer [1]. Among the major constituents of the phosphorylation machinery, which include kinases, phosphatases and phospho-binding proteins, there are a number of prominent oncogenes (FIGURE 1).

Identifying and quantifying the enormous amount of phosphorylation events, which occur both under physiological and pathological conditions, requires high-throughput techniques. The advance of mass spectrometry (MS) technology has brought with it the possibility to carry out large-scale studies of proteins and their PTMs [1].

However, biologically relevant phosphoproteins are often of low abundance and phosphorylation events are commonly of low stoichiometry. Consequently, phosphoproteome analysis by MS is challenged with identifying phosphopeptides among the ‘background noise’ of nonphosphorylated peptides. Moreover, chemical properties of phosphopeptides make them prone to get lost during sample preparation and chromatography and result in neutral losses in Collision Induced Dissociation (CID) fragmentation methods [2]. Finally, the precise localization of phosphorylated amino acids in peptides is in itself a difficult task [3–5].

Many of these challenges have been tackled by refining laboratory protocols and advancing instrumentation. Phosphoproteomics has become common practice in many laboratories, allowing analyses of vast numbers of phosphosites on a daily basis. The field is now facing the task of taking the next step and moving into the analysis of tissue samples, promising the *in vivo* portrayal of phosphorylation-mediated phenotypes, as well as identification

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.



**Figure 1. The phosphorylation machinery & how it can be analyzed with MS-based techniques.**

**(A)** The phosphorylation machinery is comprised of modification-conferring enzyme, kinases (writer), modification-removing enzymes, phosphatases (eraser) and modification interacting factors, phospho-binding domain containing proteins (readers). Those can be exploited for (MS-based) phosphoproteomics studies on multiple levels. **(B)** Different workflows are common for phosphoproteomics analyses. Few are studying phosphorylation on whole proteins (top-down APPROACH). Other studies use phosphoprotein pull-down (e.g., for phosphotyrosine enrichment), followed by enzymatic digestion. Most commonly, enzymatic digestion precedes phospho-enrichment by multiple different strategies. Those strategies are followed by LC-MS/MS analyses.

LC: Liquid-chromatography.

of novel biomarkers and therapeutic targets. The clinical application of phosphoproteomics demands the method to become faster and more user-friendly, without the loss of robustness or reproducibility. For both cell line-based and tissue-specific phosphoproteomics, a distinct challenge lies in distinguishing regulatory phosphorylation sites, which are biologically meaningful, from stochastically occurring phosphorylation events that are part of biological noise.

This review describes different technological strategies to identify phosphorylation sites with high accuracy, without significant loss of analysis speed and reproducibility in tissues and cells. Moreover, computational tools for analysis, integration and interpretation of phosphorylation events are discussed here.

### Tandem mass spectrometry of phosphopeptides

For efficient phosphopeptide sequencing and site localization, different ion fragmentation methods are used, which rely on different physical principles. In ion trap CID, peptides are fragmented by collisions with helium gas, a low-energy resonance excitation process favoring loss of the most labile bonds in proteins. Conversely, in electron transfer dissociation (ETD), anions transfer electrons to the peptides, cause release of a hydrogen radical [6] and nonergodic fragmentation without prior internal redistribution of energy. Here, cleavage of the backbone N-C $\alpha$  bond is the predominant fragmentation event. In beam-type CID or Higher Energy Collision Dissociation (HCD) on orbitrap-type instruments, higher energy levels are being deposited into the precursor peptides that are fragmented by collisions with nitrogen gas molecules [7].

Especially for ion-trap CID, loss of labile modifications is commonly observed. Phosphorylation is lost from serine and threonine residues (constituting the main pool of phosphorylated amino acid residues) as a so-called neutral loss of phosphoric acid (H $_3$ PO $_4$ ), corresponding to a net mass loss of 98 Da. The dominant neutral losses in ion trap CID type of fragmentation, at the

expense of peptide backbone fragmentation, can interfere with proper sequence assignment. This effect is largely overcome in HCD-based fragmentation. Here, phosphopeptides show less dominant neutral loss for  $y$  ion series, due to consecutive fragmentation [7,8].

Notwithstanding the development of fragmentation methods that exploit neutral losses in  $MS^3$  scanning or multistage activation [9], the main trend in the field is to use HCD on orbitrap-type instruments. Despite this tendency, different studies state better performance of low-resolution CID when compared with high-resolution HCD for phosphoproteomics analyses [10].

Neutral losses are less prevalent for ETD-based fragmentation. A main shortcoming of this method, however, lies in lower fragmentation efficiency – especially for doubly charged peptides – and subsequent decrease in sensitivity and enhanced analysis times [9]. Marx *et al.* compared the performance of HCD and ETD with an orbitrap readout on a library of synthetic (phospho)peptides. Surprisingly, they did not find better performance of ETD in terms of (phospho)peptide identification at higher charge states, as had been suggested previously [11,12]. Nevertheless, they found the techniques to be orthogonal in peptide identification and ETD to perform better for phosphorylation site localization compared with HCD [12].

Combination of different fragmentation methods is prone to yield higher coverage and complementary data. Decision-tree-based combination of ion trap CID and ETD [13,14] or ETD and HCD in the same collisional step (coined EThcD) has been shown to improve sequence coverage and phosphorylation site localization for both top-down [15] and bottom-up approaches [13,16], discussed under the following section.

The combination of different fragmentation techniques and readout modes (as possible for the Fusion and Fusion Lumos instruments from Thermo Scientific) could improve coverage and sensitivity in future studies. For example, combination of HCD or EThcD with readout of fragments in the iontrap should have the benefit of combining the most efficient fragmentation methods with the higher sensitivity of ion trap detection.

Ion mobility-based techniques, in which gas-phase ions are separated according to their mobilities in an electrical field, can serve as a complementary prefractionation technique for MS-based phosphoproteomics. Preselection for multiply charged ions can minimize interfering singly charged species and thus increase dynamic range, signal-to-noise-ratio, sensitivity and speed of MS-based analysis [17–19]. In field asymmetric ion mobility spectrometer (FAIMS) (or differential ion mobility), ions are separated in collisional cross sections in an asymmetric electric field. For phosphoproteomics, FAIMS can be useful to distinguish isomeric phosphopeptides, as shown by Creese *et al.* who found enhanced performance of liquid-chromatography (LC)-FAIMS-MS/MS over LC-MS/MS in the identification of isomeric phosphopeptides from a phosphopeptide library [20]. Also for complex samples, a combination of FAIMS with decision-tree-based acquisition scheme was shown to yield 50% higher identification rates of unique phosphopeptides [14]. Similarly, application of traveling wave ion mobility spectrometry

could improve duty cycle and sensitivity on a TOF-MS analyzer, leading to improved phosphopeptide detection and quantitation [19].

The powerful combination of ion mobility and tandem MS has the potential to significantly increase phosphopeptide coverage and will likely become an integrated part of phosphoproteomics workflows in the coming years.

## Sample preparation methods

### Phosphopeptide digestion & lysis strategies

Due to the substoichiometric nature of phosphorylation sites, phosphopeptides are commonly much lower in abundance than their nonphosphorylated counterparts. This issue demands for high amounts of starting material in phosphoproteomics experiments, to retain enough material after specific phosphopeptide enrichment techniques. Those usually leave only a small fraction of the original sample [21–23]. In general, cell lysis protocols used for phosphoproteomics are similar to those used in expression proteomics but include inhibitors of kinases and phosphatases, to block activity of those enzymes [1]. Recent studies implicated lysis with boiling guanidine hydrochloride, as a valid alternative for phosphoproteomics analyses.

Most (phospho)proteomics studies are performed in the so-called bottom-up fashion, with only a few examples of applications of top-down phosphoproteomics strategies [15,24]. Proteins are digested prior to MS-based analysis, primarily relying on tryptic digestion. This peptide-centric approach brings with it a few pitfalls that limit the discovery power of proteomics. The stochastic nature of peptide sequencing by data-dependent acquisition (DDA) (DDA – the prevailing mode of data collection in which a fixed number of precursor ions from a full MS scan are selected using predetermined rules and subjected to MS/MS analysis [25]) – innately hampers reproducibility of bottom-up proteomics studies. This is especially true for analyses of PTMs, whose substoichiometric quantities further decrease the chance to consistently sequence the same modified peptides between biological replicates. Bottom-up phosphoproteomics is further limited by the relatively short amino acid sequence of tryptic peptides (typically around 14 amino acids [26]). The short peptide length prevents the study of connectivity of phosphorylation sites or of crosstalk between phosphorylation sites and other PTMs, which occur within broader distances and hinders assignment of phosphorylation sites to specific splice variants of proteins.

Nevertheless, bottom-up strategies are still the commonly used practice in (phospho)proteomics as sequence coverage, and thereby site localization in top-down analyses is impaired by insufficient fragmentation, especially of larger proteins, as well as problems relating to chromatographic separation of intact proteins [9,15]. A few recent studies report progress in targeted and large-scale analyses of (phospho)proteins in a top-down fashion [15,24]. Tran *et al.* used a four-dimensional platform combining isoelectric focusing with gel-eluted liquid fraction entrapment electrophoresis and LC-MS for the analysis of around 1000 proteins, resulting in 3000 PTM-modified

species including 645 phosphorylation sites [24]. The complexity of this workflow, however, makes routine use of the technique unlikely. Optimization strategies in terms of fragmentation are also being implemented, as shown by Brunner *et al.* who analyzed the phosphorylation of the cell cycle kinase-target protein Bora, using different single or hybrid fragmentation methods [15]. Such targeted top-down approaches mapping phosphorylation on selected proteins will likely become common practice in the years to come.

Trypsin is a very specific protease that cleaves the peptide bond after lysine and arginine [27] and yields peptides, which are generally favorable in size for MS-based analysis [26]. Tryptic peptides are suitable targets for CID/HCD fragmentation as they are predominantly doubly charged and protonated in gas-phase on both N-term and C-term. Trypsin is by far the most popular protease used in all areas of bottom-up proteomics including phosphoproteomics, despite the fact that some tryptic peptides are too small or too large for MS-based analysis [28]. A number of studies suggest tryptic digestion to be optimal for phosphoproteome analysis [29,30]. In a study from 2014, Dickhut *et al.* indicated that the phosphorylated amino acid itself can impair tryptic cleavage, with up to 10 fold decrease in cleavage efficiency of phosphopeptides if compared with unmodified peptides [29]. The authors showed more complete cleavage of phosphorylated peptides using enhanced enzyme concentrations. However, the cost of trypsin can be a relevant factor to consider in large-scale analyses using high enzyme:protein ratios. The use of alternative peptidases such as chymotrypsin, Glu-C, AspN, LysC or subtilin can complement tryptic digests with orthogonal data of proteome regions, which are inaccessible to trypsin [1,2,28]. Nevertheless, the repressive effect of nearby phosphorylation on cleavage efficiency also seems to be present for other proteases than trypsin [31].

Despite the obvious advantages, which have made trypsin the protease of choice in proteomics and phosphoproteomics research, it is clear that for comprehensive analysis of the phosphoproteome, there is a need to use additional strategies that can cover the 'phosphorylation site space' to which trypsin is blind. Accordingly, the use of complementary proteases [31] and top-down strategies [24] will expectedly become an integral part of phosphoproteomics workflows in the years to come.

### Phosphopeptide enrichment & fractionation strategies

Phosphoproteomics analyses classically involve an enrichment step to eliminate the predominant pool of nonphosphorylated peptides from digests of whole cell lysates. Different strategies for phosphopeptide enrichment include antibody-based affinity capture, chemical derivatization of the phospho-amino acid, metal ion-based affinity capture and ion exchange chromatography [9,32].

Tyrosine phosphorylation is about 3000 fold lower in its cellular abundance compared with serine and threonine phosphorylation [33], and MS analysis is usually performed after phosphotyrosine-specific antibody-based enrichment. While previous studies have reported high amounts of starting

material (>10 mg) being necessary for specific phosphotyrosine enrichment, a recent study by the Heck laboratory yielded over 1000 unique tyrosine phosphorylated peptides from only 4 mg of starting material [34].

Immobilized Metal-Ion Affinity Chromatography (IMAC), usually Fe<sup>3+</sup>-IMAC, and Metal Oxide Affinity Chromatography, mainly titanium dioxide (TiO<sub>2</sub>) chromatography are the most frequently used enrichment techniques for S/T-phosphorylated peptides. Both techniques are well established in a great number of laboratories and can robustly yield high numbers of phosphorylation sites by LC-MS/MS analysis on routine basis [5,21–23]. Many different optimization strategies have been published and reviewed over the past decade [9,32]. The majority of those have been small incremental advances describing the use of different binding and washing conditions, which improve the number of identified phosphorylation sites. Combination of the two techniques, as well as their integration with fractionation methods can further enhance numbers of identified phosphopeptides [23,35–38].

IMAC strategies are commonly less specific than metal-oxide-based workflows as they lead to a stronger enrichment of nonphosphorylated acidic peptides. A prominent innovation in the last couple of years was the development of IMAC materials with different chelation matrices. Zhou *et al.* introduced an IMAC-Ti<sup>4+</sup> resin for metal chelation, which uses a phosphonate moiety for immobilization, and Ti<sup>4+</sup> as the metal cation [39,40]. It resembles TiO<sub>2</sub>-based enrichment in having high tolerance toward acidic buffers and low nonspecific binding of acidic peptides [11] and was successfully implemented to different high-throughput phosphoproteome analyses in cell lines and tissues [13,41].

Recently, Fe-IMAC in an HPLC-based format was found to outperform Ti-IMAC and TiO<sub>2</sub> in direct comparison [42]. Interestingly, this study challenged the generally accepted view of IMAC and Metal Oxide Affinity Chromatography-based techniques being orthogonal in terms of phosphopeptide identification, a striking finding that will require further validation in the field.

In general, the number of identified phosphorylation sites strongly depends on sample amounts, chromatography, MS instrumentation and analysis pipelines, rendering the direct comparison of the phosphorylation site identifications from different laboratories difficult (TABLES 1–3).

A number of alternative enrichment strategies have been developed in recent years, including polymer-based metal ion affinity capture (PolyMAC) [43,44], Phos-tag [45], as well as precipitation-based enrichment strategies, such as those based on lanthanide ions [46]. Those techniques are further complemented with targeted strategies enriching specific subsets of the phosphoproteome, such as motif-specific antibodies [47], or 'fishing' with phospho-binding domains such as phosphotyrosine binding SH2 domains [48] or 14–3–3 motifs [49].

Phosphopeptide enrichment is commonly linked with fractionation strategies to reduce sample complexity. Strong cation exchange (SCX)-based chromatography has been the most

**Table 1. Quantitative phosphoproteomics in cell lines.**

Study (year)	Biological context	Enrichment method	Quantitation	Identified sites	Ref.
Olsen <i>et al.</i> (2006)	HeLa cells EGF-stimulated	TiO <sub>2</sub> and SCX	SILAC	6600 phosphorylation sites on 2244 proteins	[5]
Nagaraj <i>et al.</i> (2010)	HeLa	SCX/TiO <sub>2</sub>		9668 phosphorylation sites	[8]
Rigbolt <i>et al.</i> (2011)	Human stem cells	SCX/TiO <sub>2</sub> and others	SILAC	23,522 phosphorylation sites, 6521 proteins	[52]
Beli <i>et al.</i> (2012)	Human osteosarcoma cells (U2OS) treated with etoposide or ionizing radiation	SCX/TiO <sub>2</sub>	SILAC	1470 phosphorylation sites (additionally 1796 lysine acetylation sites)	[74]
Mertins <i>et al.</i> (2012)	HeLa cells EGF-stimulated	SCX/IMAC	iTRAQ & mTRAQ	12,129 phosphopeptides on 2699 phosphoproteins (iTRAQ); 4448 phosphopeptides on 1597 phosphoproteins	[77]
Engholm-Keller (2012)	Rat cell line (INS-1)	'TiSH': TiO <sub>2</sub> /sequential elution from IMAC/HILIC		6600 unique phosphopeptides	[35]
Batth <i>et al.</i> (2014)	Mouse 3T3 cells	off-line high-pH chromatography TiO <sub>2</sub>		30,000 unique phosphopeptide variants	[37]
Bennetzen <i>et al.</i> (2010)	GM00130 cells DNA damage-exposed	ERLIC and TiO <sub>2</sub>	SILAC	5204 phosphorylation sites	[73]
Bensimon <i>et al.</i> (2010)	G361 human melanoma cell line DNA damage-exposed	TiO <sub>2</sub>	Label-free	2871 phosphorylation sites on 1099 proteins	[75]
Mertins <i>et al.</i> (2013)	Jurkat cells (human T-cells)	Fe <sup>3+</sup> -IMAC		20,800 phosphorylation sites	[35]
Sawney <i>et al.</i> (2013)	Yeast	SCX	Label-free	2100 phosphorylation cells, 466 proteins	[107]
FrancaVilla <i>et al.</i> (2013)	HeLa cells FGF-stimulated	Antiphosphotyrosine beads	SILAC	1212 tyrosine phosphorylated peptides	[71]
Kumlerl <i>et al.</i> (2014)	Yeast (rapamycin-treated)	SCX/TiO <sub>2</sub>	SILAC	8961 phosphorylation sites, 3590 proteins	[136]
de Graaf <i>et al.</i> (2014)	Jurkat cells (human T-cells)	Ti(4+)-IMAC and phosphotyrosine antibody enrichment	Label-free	16,200 phosphorylation sites	[13]
Ruprecht <i>et al.</i> (2015)	Human epidermoid A341 cells	Fe-IMAC in an HPLC format Combined with SCX	Label-free	14,00 unique phosphopeptides	[42]
Giansanti <i>et al.</i> (2015)	Jurkat T-cells	Ti <sup>4+</sup> -IMAC	Label-free	18,430 unique phosphosites	[31]

Cell studies of phosphoproteomics listing: Biological context of the experiments, Phosphopeptide enrichment method, quantitation method, number of identified sites and reference.

ERLIC: Electrostatic repulsion hydrophilic interaction chromatography; HAMMOC: Hydroxyl acid-modified metal oxide chromatography; iTRAQ: Isobaric tags for relative and absolute quantitation; SCX: Strong cation exchange; TiO<sub>2</sub>: Titanium dioxide.

widely applied off-line fractionation method to date [5,23]. However, also hydrophilic interaction chromatography (HILIC), electrostatic repulsion-hydrophilic interaction chromatography and pH reversed-phase chromatography have been

successfully applied in large-scale phosphoproteomics studies [35,37,38,50]. Notably, due to its high phosphopeptide separation power, high pH reversed-phase chromatography is becoming popular as an alternative to SCX-based fractionation.

**Table 2. Stoichiometry.**

Study (year)	Cell line	Enrichment method	Quantitation	Stoichiometry estimation	Identified sites	Ref.
Olsen <i>et al.</i> (2010)	HeLa (cell cycle progression)	SCX/TiO <sub>2</sub>	SILAC	Data-dependent stoichiometry estimation	20,443 phosphorylation sites, 6027 proteins	[22]
Wu <i>et al.</i> (2011)	Yeast	HILIC	Chemical labeling with stable isotopes	Phosphatase treatment	5033 phosphorylation sites	[79]
Sharma <i>et al.</i> (2014)	HeLa (Cell cycle progression and EGF stimulation)	SCX/TiO <sub>2</sub>	Label-free	Data-dependent stoichiometry estimation	50,000 phosphorylation sites	[23]
Tsai <i>et al.</i> (2015)	Gefitinib-sensitive and resistant human lung adenocarcinoma cell line PC9, Raji human B cell line	IMAC purification (Followed by kinase reaction)	Isotopic tagging with dimethyl labeling	Phosphatase treatment in combination with kinase assay	1000 phosphorylation sites	[80]

Phosphoproteomics analyses including stoichiometry measurements: Biological context of the experiments, Phosphopeptide enrichment method, quantitation method, stoichiometry measurement method, number of identified sites and reference. SCX: Strong cation exchange; TiO<sub>2</sub>: Titanium dioxide

The powerful linkage of phosphopeptide enrichment and fractionation can yield impressive coverage of the phosphoproteome. This was shown in a recent study of the Mann laboratory, in which they identified more than 50,000 distinct phosphopeptides in the human HeLa cell line, the most popular cellular system used in phosphoproteomics to date (FIGURE 2). In this study, the authors identified more than three-quarters of the HeLa proteome as target for phosphorylation. Those included base-level phosphorylation events and phosphorylation

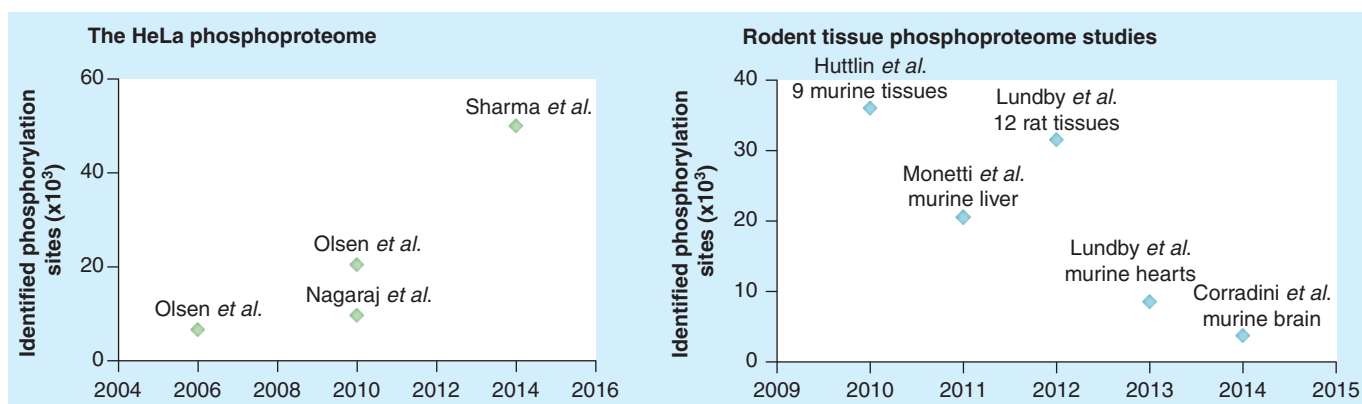
sites responding to EGF treatment, mitotic arrest and pervanadate treatment [23].

Compared with this and other deep analyses of the phosphoproteome [21,22,51,52], a number of studies have shown the benefit of enrichment without prior or postfractionation and 'single-shot' MS analysis [53]. Using equal amounts of starting material, Kettenbach *et al.* could yield comparable numbers between a combination of SCX and TiO<sub>2</sub> enrichment and single TiO<sub>2</sub> enrichment. The authors highlighted the increase in

**Table 3. Tissue phosphoproteomics.**

Study (year)	Tissue	Enrichment method	Quantitation	Identified sites	Ref.
Monetti <i>et al.</i> (2011)	Murine liver	SCX/TiO <sub>2</sub>	Spike-in SILAC	up to 20,491 phosphorylation sites	[65]
Corradini <i>et al.</i> (2014)	Murine brain	Ti <sup>4+</sup> -IMAC	Manual inspection	3690 identified 3271 quantified	[67]
Lundby <i>et al.</i> (2012)	14 rat organs and tissues	TiO <sub>2</sub>	Label-free	Total 31,480 phosphorylation sites, 7280 proteins	[55]
Huttlin <i>et al.</i> (2010)	9 murine tissues	SCX/ Fe <sup>3+</sup> -IMAC	Label-free	35,965 phosphorylation sites, 6296 proteins	[21]
Lundby <i>et al.</i> (2013)	Murine hearts	TiO <sub>2</sub>	Label-free	8518 phosphorylation sites, 4246 proteins	[64]
Narumi <i>et al.</i> (2012)	Human breast cancer tissues	Fe <sup>3+</sup> -IMAC	iTRAQ SRM-based validation	8309 phosphorylation sites on 3401 proteins 19 validated by SRM	[68]
Wakabayashi <i>et al.</i> (2013)	Formaline-fixed, paraffin-embedded and fresh murine liver	HAMMOCC	Chemical labeling based on reductive dimethylation	1090 phosphopeptides	[62]

Large-scale tissue-specific phosphoproteomics studies. Cell studies of phosphoproteomics listing: Biological context of the experiments, phosphopeptide enrichment method, quantitation method, number of identified sites and reference. HAMMOCC: Hydroxyl acid-modified metal oxide chromatography; iTRAQ: Isobaric tags for relative and absolute quantitation; SCX: Strong cation exchange; TiO<sub>2</sub>: Titanium dioxide.



**Figure 2. Phosphosite identifications over time.** Plot shows increase in identification of phosphorylation sites in HeLa cells and different rodent tissues over the last 9 years.

reproducibility and decrease in analysis time achieved by eliminating the prefractionation step [53]. Moreover, single shot methods are beneficial for the scalability of phosphoproteomics experiments.

Different groups reported identifications in the range between 1000 and 10,000 phosphorylation sites from sample amounts between 1 and 400  $\mu\text{g}$ , including both single-shot and fractionation-based methods [9,35,50,53,54]. The Ishihama group could, for example, identify 1000 phosphorylation sites from as little as 10,000 HeLa cells, by using a combination of hydroxyl acid-modified metal oxide chromatography and a miniaturized LC column of 25  $\mu\text{m}$  internal diameter [54].

High coverage phosphoproteomes are unquestionably relevant to clarify the 'complete picture' of single cellular states or responses to relatively small numbers of stimuli. The data assembled in those studies provide invaluable compendia of possible phosphorylation sites. Those are far from comprehensive to date, as shown by the continuous identification of previously undescribed phosphorylation sites.

In contrast, faster, 'single-shot' methods will be vital for clinical application of phosphoproteomics, where changes in phospho-signatures rather than analysis of individual phosphorylation sites are the goal. Application of such 'minute amount of input material' protocols will expand our view of phosphorylation-mediated processes in lowly abundant primary cell types, such as stem cells. Single-shot methods, which robustly identify biologically relevant phosphorylation sites, could furthermore bridge the gap between targeted and discovery phosphoproteomics studies and support establishing phosphoproteomics as a routine technique in a broader number of laboratories.

### Challenges in tissue phosphoproteomics

Despite the wealth of phosphoproteomics studies analyzing a great number of biological phenotypes in cell lines, phosphoproteomics investigations of tissues are still relatively sparse.

Next to the challenges common to phosphoproteomics analyses of cell lines, tissue phosphoproteomics are faced with a number of distinctive limitations. The first of those relates to

sample availability, especially for analysis of small organs in rodent models or prognostic tissue biopsy samples from patients. Thus, it can be difficult to yield adequate sample amounts for efficient phosphopeptide enrichment. Another factor is sample heterogeneity due to biological variation, as well as the presence of blood. The latter can be reduced by perfusion during euthanization, when studying rodent models [51,55]. Heterogeneity among tissue cell types or among different tumor cell populations can potentially be addressed by tissue microdissection prior to MS analysis [56].

Sample collection conditions inevitably introduce a bias for both cell line and tissue-based phosphoproteomics [57–60]. For phosphoproteomics immediate inactivation of cellular kinases and phosphatases is a prerequisite, but this can be especially difficult for collection of clinical samples of human patients. In the worst case, artifacts related to sample collection and preparation conditions could lead the researcher to assume false regulation of phosphorylation pathways, as discussed below.

Recent studies have analyzed the effect of ischemic events on the proteome, total phosphoproteome and the tyrosine phosphoproteome of tumor samples [59,60]. While those studies found the proteome largely unchanged, up to one fourth of the total phosphoproteome and half of the tyrosine phosphoproteome were significantly affected by ischemic events [59,60]. Moreover, proteins, which showed altered phosphorylation after ischemia, were often linked to cancer-relevant pathways. This hints that deviations of those pathways potentially inferred from tissue phosphoproteomics studies could in reality result from ischemic events [59,60]. It is, therefore, important to make use of rapid preparation methods that 'freeze' the *in vivo* state of the phosphoproteome, for example, by protein denaturation using heat-inactivation of fresh or snap frozen tissue [51,61].

Interestingly, a recent report by the Ishihama group suggested the potential to perform quantitative phosphoproteomics analysis on paraffin-embedded murine liver tissue. A specific sample preparation protocol was applied, including sample heating to break formalin-induced protein crosslinks and paraffin removal by ethyl acetate. Subsequent application of the hydroxyl acid-modified metal oxide chromatography method

combined with miniaturized LC setup yielded identification of 1090 phosphopeptides from a single paraffin section (roughly 25 µg of protein), with the 'paraffin'- phosphoproteome being similar to that of fresh liver tissue samples [62].

Different groups catalogued tissue-specific phosphorylation in model organisms, such as mouse, rat or drosophila embryos [21,51,63]. Among the more than 30,000 phosphorylation sites that were found in the analyses of mouse and rat phosphoproteomes, the authors found roughly 50% of phosphorylation events to be tissue specific, with especially high numbers in brain and testis [21,51]. Huttlin *et al.* found that the most abundant phosphorylation motif was proline-directed, comprising around 30% of all phosphorylation sites [21].

While those studies encompass important repositories of tissue-specific and global phosphorylation events, they do not analyze the reaction to perturbations. Different reports now address *in vivo* responses to stimuli, such as response to β-blockers and activators in murine hearts [64], insulin signaling in mouse liver [65], antagonistic antibodies in xenograft models of human breast cancer cell lines [66] and the phosphorylation differences between wild-type and guanosine monophosphate-dependent protein kinase CGKI knockout mice [67], to name a few.

In a report from 2012, Narumi *et al.* indicated the potential for selected reaction monitoring (SRM), a highly sensitive, targeted MS/MS method, to corroborate biomarkers identified in large-scale discovery phosphoproteomics. The authors performed SRM-based validation of 19 phosphopeptides, which were differentially phosphorylated between tissue samples of patients with breast cancer from high- and low-risk recurrence groups [68].

Despite the increasing number of studies applying phosphoproteomics to tissues, a general lack of highly standardized protocols equivalent to those that exist for the analysis of cell lines, conditions of sample collection and issues relating to phosphosite quantitation still constitute bottle-necks for tissue phosphoproteomics. The recent description of possibilities to perform phosphoproteomics on paraffin-embedded tissues might be intricately relevant for clinical applications of phosphoproteomics [62].

### Quantitative phosphoproteomics in cell lines and tissues

Most phosphoproteomics analyses are targeted toward finding alterations in biological phenotypes, for example, between treatment conditions, diseased versus healthy cells, tissues or organisms or developmental stages. For this quantitative comparison of biological states, a number of techniques are available, including metabolic or isobaric labeling, as well as label-free techniques. Stable isotope labeling strategies include SILAC, dimethyl labeling and the use of isobaric tagging reagents such as isobaric tags for relative and absolute quantitation (iTRAQ) and the tandem mass tagging (TMT) approach. The MS1 level-based quantitation strategies SILAC and dimethyl labeling are restrained by the availability of different labels, as complexity of spectra is a limiting factor. For isobaric labels, quantification of peptides is achieved on the MS2 level, allowing for the

use of more labels, often 4, 8 or even 10 but also multiplexing approaches [69,70].

SILAC-based phosphoproteomics is a robust and well-established methodology, which has been extensively applied to study dynamic signaling responses in cell line models. Those studies included responses to growth factors [5,71] and kinase inhibitors [72], cell cycle stage [22], DNA damage responses [47,73–76] and ES cell differentiation [52].

Advantages and disadvantages of different strategies relate to handling, reproducibility, efficiency and price. While SILAC labeling is very robust and comparably inexpensive for phosphoproteomics, iTRAQ/TMT labeling is much more expensive. Due to mixing at later stage, iTRAQ/TMT strategies are more prone to artifacts introduced during sample preparation, which hampers accurate quantitation.

For SILAC or dimethyl labeling, the precursor signal is split, increasing the full-scan complexity and limiting the dynamic range. Conversely, iTRAQ/TMT labeling has an additive effect on precursor intensities. In a direct comparison of chemically identical nonisobaric, mass differential tags for relative and absolute quantification and iTRAQ labels, Mertins *et al.* found that iTRAQ led to the identification of three times more phosphopeptides [77], but at the expense of compressed phosphopeptide ratios due to coisolation and MS/MS analysis of coeluting phosphopeptides of similar mass-to-charge.

Similar to isobaric labeling, label-free techniques are vulnerable to sample preparation-induced biases. In addition, the lack of a mixing step and requirement for a higher number of replicates can increase the use of instrument time substantially. A specific challenge for label-free quantitation of PTMs lies in the fact that modified peptides have to be quantified individually, as opposed to proteomics analyses, where multiple peptides contribute to the quantitation of a protein. Missing quantitation points in individual samples can thus hamper label-free analyses. This issue is typically addressed by a so-called match between run approach, where peptide intensities are transferred between replicates by matching the accurate mass and retention time of peptides identified in one of the experiments. However, it is important to keep in mind that this approach is challenging in label-free phosphoproteomics dataset because differentially phosphorylated versions of the same peptide can have the same mass and often similar elution profiles and retention times.

Nevertheless, rigid protocols for sample preparation, good and reproducible chromatography and MS analysis, and robust analyses methods [78] make application of label-free techniques for both tissue and cell line-based phosphoproteomics feasible [23,64]. For future use in clinical phosphoproteomics, accurate label-free strategies are highly relevant, as costly isobaric labeling techniques might not be applicable for routine use.

### Determining phosphorylation occupancy

Next to identifying changes in phosphorylation levels between different biological conditions, the determination of the relative stoichiometry of phosphorylated forms of proteins is an important aspect.



Perceptibly, changing ratios of phosphorylated peptides (for example, in response to drug exposure or during cell cycle progression) might have drastically different biological effects in cases where only low amounts of the total protein are phosphorylated as opposed to cases where phosphorylated forms show very high occupancy. However, identifying phosphorylation site occupancy is not a trivial task, and only a small number of studies have dealt with measuring occupancy. In 2010, Olsen *et al.* measured occupancy of various phosphorylation sites throughout the cell cycle, by global quantitative (phospho)proteomics. The authors used SILAC ratios of protein, phosphopeptide and unmodified peptide, to calculate occupancy of phosphorylation sites. They found around 50% of the sites on which they could perform stoichiometry calculations to be phosphorylated with occupancy of greater than 70% during mitosis. This method is very practical as it calculates stoichiometry directly from the data, without introducing further experimental steps, but occupancy can only be calculated for phosphorylation sites, which show altered SILAC ratios. Moreover, the technique is only applicable for singly phosphorylated peptides, as multiply phosphorylated peptides would necessitate more complex methods of calculation [22].

The same method was applied for other combined high-coverage analyses of proteome and phosphoproteome. Sharma *et al.* performed ultra-deep analysis of the HeLa phosphoproteome after stimulation with the growth factor EGF or release into mitosis, using a label-free approach and calculated fractional occupancy on 7620 sites. They could confirm high stoichiometry of mitotic phosphorylation sites, as opposed to low phospho-stoichiometry of EGF-stimulated and control cells (<25%). Measuring the fractional occupancy of tyrosine phosphorylation for the first time, they showed that tyrosine phosphorylation, in contrast to serine and threonine phosphorylation, had higher occupancy in EGF-treated cells when compared with control or mitotic cells [23].

In 2011, a study by Wu *et al.* implemented a method based on combination of phosphatase treatment and stable isotope labeling, to determine absolute protein phosphorylation stoichiometry in yeast [79]. They subjected samples to phosphatase or mock treatment and subsequent chemical labeling. The authors used a list of 5033 phosphorylation sites described in the literature and determined the occupancy of their phosphorylation. Only few sites showed very high occupancy, while more than half of the investigated sites showed occupancy lower than 30%. GO-term analysis revealed that phosphosites with high occupancies related to chromatin silencing and cytokinesis during cell cycle, the latter being in line with the studies by Olsen and Sharma *et al.* [22,23]. While this method is sophisticated for identifying phosphorylation stoichiometry, it lacks the potential for new discovery, as it relies on prior knowledge. Moreover, also here multiphosphorylated peptides constitute a problem for occupancy determination.

In a recent elegant study, Tsai *et al.* implemented a kinase motif-targeting approach for analysis of phosphorylation site occupancy of specific kinase substrates. They analyzed a single state human proteome, by combining dephosphorylation with

isotope tagging and enzymatic kinase reactions [80]. This strategy allowed them to estimate phosphorylation site occupancy of >1000 phosphorylation sites that were targets of ERK2, CK2 and EGFR.

Next to identifying the relative stoichiometry of phosphorylation, identifying absolute levels of phosphorylated peptides can be of interest [81]. Absolute (phospho)peptide quantification generally relies on spike-in of synthetic, isotopically labeled standard peptides, which can then be used to quantify absolute amounts of the correspondent peptide present in the sample. Combination of heavy-labeled standard peptides with isobaric tagging has recently been applied to analyze quantitative differences in phosphorylation of the EGFR in response to different ligands [82].

It should be noted that the described effect of phosphorylation sites on protease-based digestion [29,31] might bias stoichiometry estimations. If indeed some phosphopeptides show 10-fold lower tryptic digestion efficiency, this could lead to strongly underestimated occupancy calculations of those phosphopeptides [2,29].

MS is still an analytic technique that determines phosphorylation events on the population and not yet on single cell level. Thus, heterogeneity in a population of cells cannot be reflected. As site occupancy changes throughout the cell cycle, in asynchronous cell populations [79] with only a few cells in mitosis, bulk occupancy values could be misleading. Orthogonal technologies such as high-throughput flow cytometry or microscopy might in many cases be necessary to reveal phosphorylation stoichiometry on the level of individual cells.

### Finding biologically relevant phosphosites

To study the functional consequence of identified phosphorylation sites, it is crucial to localize the phosphorylation site within the peptide with single amino acid precision. High-quality MS/MS data are a prerequisite for such exact site localization. Protein and peptide identification is based on database search tools that match peptide fragmentation patterns to a database of known protein sequences [83]. However, those methods are usually not optimized for phosphorylation site identification, as they do not indicate confidence of site localization. Most peptides contain more than one phospho-modifiable amino acid, complicating the precise localization of the phosphosite. Phosphosite localization, in general, makes use of two related probability-based scoring algorithms; the A-score by Beausoleil *et al.* [3] and the PTM Score by Olsen *et al.* [4,5], as well as different scores derived from those. In a study comparing three different scoring algorithms on a library of synthetic phosphopeptides, Marx *et al.* found similar performance for the three popular scores: Mascot Delta score, PTM-score (implemented with Max-Quant and its Andromeda algorithm) and the phospho-RS score (embedded into proteome discoverer) [12].

Among more than tens of thousands of phosphorylation events identified in a routine phosphoproteomics experiment, nowadays [5,13,21,23,37,52,77] only a subset is relevant for specific cellular signaling processes. Many other phosphorylation events

are either involved in structural processes or can be considered 'biological noise' [84]. One reason for this 'noisiness' might be that kinases, similar to other cellular enzymes can make errors in substrate recognition and phosphorylation [2,85]. Moreover, occupancy of phosphorylation of certain residues might have to reach a certain threshold to perform a biologically meaningful function, indicating that the phosphorylation at the time of analysis is not vital for observed phenotype(s).

Different studies used MS-based phosphoproteomics as the means or basis for hypothesis-driven biological experiments, as shown in recent comprehensive MS-based analyses of RTKs signaling networks [71,86–89].

Various bioinformatics strategies can help providing insights into the potential functionality of phosphorylation sites identified in nonhypothesis-driven large-scale phosphoproteomics analyses. Those can be generally divided into predictive or descriptive strategies [90]. Most bioinformatics and mathematical analyses of MS-based phosphoproteomics still fall into the descriptive category, with a few studies favoring predictive models, based on phosphoproteomics datasets [90].

### Bioinformatics strategies to analyze phosphoproteomics datasets

Among the descriptive approaches both data clustering and enrichment analyses are very common. When dealing with datasets of global phosphorylation site changes spanning multiple settings, such as different time points or treatment conditions, clustering analyses can be a highly valuable tool to identify closely related phenotypes [5,71,91]. Olsen *et al.* used fuzzy *c*-means clustering to identify groups of phosphosites, which had related dynamic profiles, with identified clusters containing functionally related members [5]. Francavilla *et al.* used the same clustering approach for phosphoproteome changes that were induced by treatment with two different FGF ligands. Clusters of varying dynamic profiles differed substantially for the response to the two different ligands [71].

Next to clustering, enrichment of functional categories, for example, from the gene ontology [92] or KEGG database [93] is often applied to infer information about biological functions of groups of phosphorylation sites. Various online tools such as David [94] or InnateDB [95] allow enrichment analyses mining various databases. Care should be taken, however, regarding enrichment versus whole genome data, as the unavoidable experimental bias of MS-based experiments can skew the enrichment considerably. It is therefore recommendable to make use of the nonregulated phosphoproteome from a given experiment as the 'background' for enrichment analyses.

Both clustering and enrichment are applied to almost all bigger phosphorylation analyses nowadays and serve as the basis for hypothesis generation for biological follow-up experiments [5,22,52,71,73–75]. GO-term analyses of phosphorylation responses to two different FGF ligands indicated changes in cell migration to be specific to FGF-10, while phosphorylation responses mediated by FGF-7 were enriched for cell proliferation [71]. The authors could confirm the hypothesis based on

migration assays in breast cancer cells, and *in vivo* branching of murine lung explants. This study constitutes one of the rare examples where the hypotheses that are generated by GO-term analyses of phosphoproteomics studies are being functionally validated in both *in vitro* and *in vivo* context [71].

Mapping phosphorylation sites onto protein interaction networks is another way to gain functional knowledge about regulated phosphoproteins. The interactions reported within those databases have varying degrees of confidence as they can be based on manual curation, text mining, experimental data, evolutionary conservation and so on [96–98]. Visualizing interactions with tools such as STRING DB [96] or Cytoscape [99] can give insights into the biological context in which phosphorylation sites function. In 2010, Klammer *et al.* introduced an algorithm, which combines phosphoproteomics data with STRING-based protein interaction network information. The subextractor algorithm is based on Bayesian probabilistic modeling, which takes into account both differential regulation and network topology [100].

A great number of phosphoproteomics studies have used network-based analyses to infer biological roles for identified phosphoproteins. Beli *et al.* used STRING-based network analyses to find functional modules in DNA damage-exposed human tumor cells. They found a cluster centered on splicing and subsequently validated the recruitment to or exclusion from sites of DNA damage for the splicing-related factors PPM1G and THRAP3 [74]. Similar to GO-term analysis and clustering, network-based analyses is good starting point for hypothesis generation from large-scale phosphoproteomics studies.

A sophisticated description of biological phenotypes by the use of mathematical models, which are based on phosphoproteomics data, has been attempted in only a few studies to date [90].

Different groups aimed to use phosphoproteomics analyses of cancer cell lines for the classification of tumor types. In 2013, Casado *et al.* used regression-based modeling to classify a panel of hematological cancer cell lines and were able to distinguish sensitive and resistant cells based on patterns of their kinase activities [101]. Interestingly, they found that kinase activities indicative of resistance to certain drugs were not necessarily members of the pathway that was being targeted by the respective drug.

In 2014, Vaga *et al.* combined a set of logic models of MAPK signaling in yeast with high-time resolution measurements of phosphoproteomics and identified extensive crosstalk between two MAPK signaling pathways [102]. Another study of the HOG pathway in yeast by Kanshin *et al.* provided high-resolution temporal phosphorylation data in response to osmotic stress allowing the kinetic modeling of this phosphorylation response [91].

As indicated by those studies, a large number of datapoints is necessary for sophisticated modeling, which is most often limiting in MS-based analyses. Phosphoproteomics of model organisms such as yeast now allow for a more fine-grained

temporal or stimulus-based analysis of global phosphorylation changes [91,102] and might serve as the foundation for deriving and testing mathematical models of phospho-signaling networks.

### Analysis of sequence specific features

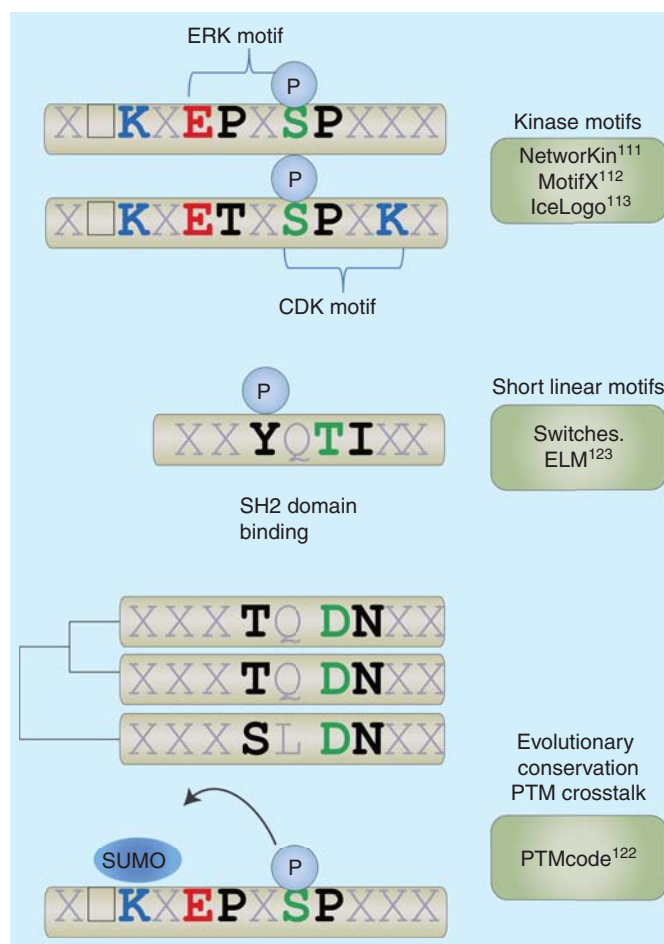
The above-described analysis strategies treat phosphorylation site changes as if there were changes in gene expression or protein abundance. This way of analysis neglects a crucial layer of information, which is provided in the phosphorylation sites and the surrounding amino acid sequence context.

A number of strategies rely on analyzing sequence information in the vicinity of phosphosites (FIGURE 3). Many of those have been implemented as web-based or standalone tools or have been incorporated into analyses software packages. Interestingly, the presence of multiple phosphorylation sites within the same peptide is rarely accounted for in bioinformatics analyses. This constitutes a conceptual problem, as different phosphorylation sites on the same peptide or protein can follow different dynamic profiles and have different functional roles. For example, a number of phosphorylation events serve as ‘priming sites’ for subsequent phosphorylation of the same protein [5,71]. Despite great experimental efforts targeted toward identifying multiphosphorylated peptides [9,35], those peptides pose a clear challenge for analysis, related to site localization ambiguity, stoichiometry estimation and bioinformatics integration.

### Analyzing evolutionary conservation

Studies of the evolutionary conservation of the localization and sequence bias of phosphorylated amino acids have yielded conflicting results [84]. Phosphorylation sites are generally assumed not to be well conserved between species [103], despite the fact that phosphorylation site numbers of proteins within signaling pathways are similar between species [104]. By functional prioritization of many MS-based PTMs studies, which was based on cross-regulatory events, domain activity and protein–protein interactions, Beltrao *et al.* analyzed their evolutionary conservation and functionality. They confirmed the assumption that phosphorylation sites with known biological functions, as well as phosphorylation sites that have been shown to be regulated in MS-based experiments, are more likely to be evolutionary conserved [105]. They, moreover, found an increase of evolutionary conservation of phosphorylation sites [105], which occur in close proximity to other PTMs, hinting toward the importance of PTM–crosstalk [106], as discussed below. A later study of co-occurring phosphorylation and ubiquitylation events in yeast confirmed higher evolutionary conservation of phosphorylation sites, which co-occur together with ubiquitylation events [107]. A high-resolution temporal analysis of phosphorylation signaling after osmotic shock in yeast further hinted toward evolutionary conservation of fast-occurring versus slow-occurring phosphorylation events [91].

When comparing different fungal species, Wu *et al.* found sites, which showed high phosphorylation occupancy to be less conserved than the ones that showed lower occupancy [79]. In



**Figure 3. Site-specific features in phosphoproteomics analyses.** The graph shows different levels of information to be gained from sequence surrounding the phosphorylated amino acid, such as kinase-motifs, evolutionary conservation, PTM-crosstalk and other regulatory elements. Tools to explore the different levels of information are highlighted. PTM: Posttranslational modification.

contrast to this, other studies reported high occupancy phosphorylation sites to show a higher evolutionary constraint [84,108]. In this respect, it is interesting to note that phosphorylation sites are five-fold less conserved between human and drosophila when compared with lysine acetylation [109]; however, the latter are generally of very low stoichiometry compared with phosphorylation [110]. This might in part be a result of lysines residues being on average more conserved than serines and threonines, even when not modified.

The low evolutionary constraint of annotated phosphorylation sites without a cellular function might be a result of their functional divergence throughout species progression. It could, however, also be related to a general lack of function, making it less evolutionary favorable to preserve those sites. Nonetheless, kinase–substrate interactions are generally predicted to be more evolutionary conserved than specific phosphorylation sites, with the potential of different sites within the substrate or

on a different subunit within the same complex being the phosphorylation target of the same kinase in different organisms [111,112]. For the prioritization of large-scale phosphorylation data, as well as for extrapolating findings from one biological system to another, those factors should be taken into consideration. Moreover, it is interesting to consider those findings, as well as the fact that a few point mutations are often sufficient to alter a PTM site [84] in the context of cancer cells. The genomes of cancer cells have high mutation rates and cancer cells often show generally aberrant signaling responses [113]. For biomarker discovery, it is thus exceedingly important to identify the biologically relevant and robust phosphorylation sites, among the jungle of nonfunctional phosphorylation events.

### Kinase–substrate relationships

While phosphoproteomics provide a wealth of data, identifying kinases responsible for specific cellular phenotypes can be a difficult task. Mapping kinase–substrate relationships is a critical step for understanding essential signaling networks and identifying pharmaceutical targets for drug discovery [114,115].

Various tools (e.g., Phospho.ELM [116], NetworKIN [117], motif-x [118], and IceLogo [119]) can enrich for kinase motifs among the regulated phosphorylation sites, hinting toward apical modulators of the observed responses. The NetworKIN algorithm links contextual information with predicted kinase motifs to identify upstream kinases, causing a higher accuracy of this algorithm [117]. A combination of motif-x, for identification of linear motifs, and NetworKIN, for the prediction of the responsible kinase, has been applied in various studies for identification of kinases likely responsible for the observed phenotypes, such as human embryonic stem cell differentiation [52,91].

However, not all kinases have favored linear sequence motifs. It seems only serine/threonine kinases but not tyrosine kinases have specific sequence motif preferences. Instead, specificity in the recognition of tyrosine phosphorylation motifs is achieved by the readers of this modification, that is, proteins with SH2 or PTB domains (FIGURE 1) [120]. Moreover, members of kinase families share the same motifs, and kinases rarely show exclusively ‘direct’ effects [121]. The sum of those factors indicates a generally highly complex relationship between kinase groups and their direct and indirect targets, obstructing inference of direct kinase–substrate relationships from phosphoproteomics data. An exception to this rule seems to be the secreted phosphoproteome that was recently shown to be almost exclusively targeted by a single kinase, Fam20C [122].

Direct kinase–substrate relationships have classically been validated using biochemical assays relying on radioactive isotope labeling of phosphorylated targets with  $^{32}\text{P}$  or  $^{33}\text{P}$ , which allow phosphorylation analysis status by radioactivity measurements.

A number of MS-based strategies can be applied to identify kinase–substrate relationships *in vitro* and *in vivo*. Those include the pharmacological inhibition of kinases [72,75,123], use of analogue-sensitive kinases [124], kinobead-technology [125] and

application of native MS for studying protein phosphorylation [126]. Moreover, *in vitro* kinase assays on whole cell extracts have been combined with MS-based phosphoproteomics to identify substrate pools of different kinases [127–131]. Those included, for example, the protein-tyrosine kinase SYK [129] or the serine/threonine ERK [130], as well as kinases of the Plk-family [131].

### Analysis of crosstalk between phosphorylation & other PTMs

Phosphorylation does usually not occur as an isolated event on a protein. Crosstalk or at least co-occurrence between multiple phosphorylation sites or between phosphorylation and other PTMs is common, and often occurs in close proximity on proteins [84,132]. This permits investigating them using classical tryptic peptide-based analyses [107].

Co-occurrence of modifications on the same protein can introduce combinatorial logic [107]; and, for example, hinder or promote occurrence of additional PTMs [106]. PTM crosstalk can regulate and fine-tune the function of key signaling hub proteins such as histone tails [133] or the tumor suppressor p53 [134].

Several groups have begun to study the crosstalk between different PTMs using MS [74,107,135,136]. Using different enrichment protocols they aimed to decipher crosstalk of ubiquitylation and phosphorylation in yeast, or the crosstalk of phosphorylation, acetylation and ubiquitylation in human cancer cell lines [74,135]. Beli *et al.* performed parallel enrichment for phosphopeptides and acetylated peptides, of DNA damage-treated human tumor cells [74]. Mertins *et al.* commenced their sequential enrichment protocol with IMAC-based phosphopeptide enrichment and continued with antibody-based enrichment of ubiquitylated proteins and acetylated peptides [135]. Swaney *et al.* isolated His-tagged ubiquitylated protein by cobalt-NTA affinity purification, successively enriched phosphopeptides from both ubiquitylated and nonubiquitylated fractions and finally identified ubiquitin sites with a di-Gly antibody [107].

Neither of the above approaches can give information about the co-occurrence of PTMs on the same protein. Implementing an additional peptide-centric approach, Swaney *et al.* used SCX-based charge-state fractionation of peptides, which were subjected to di-Gly antibodies to identify ubiquitylated phosphopeptides [107].

In cases where the ‘crosstalk’ motifs between different PTMs are described, motif analyses of the region surrounding the phosphorylation site can give hints toward the occurrence of this crosstalk in different biological contexts. Moreover, large-scale sequence analyses of phosphorylated peptides can also identify other regulatory elements, such as enrichment of degradation signals of the KEN-box and PEST region type on cell cycle-regulated (phospho)peptides [22]. An interesting repository of linear motifs, which regulate protein function, can be found in the switches. ELM resource that was introduced by van Roey *et al.* [137].

### Expert commentary

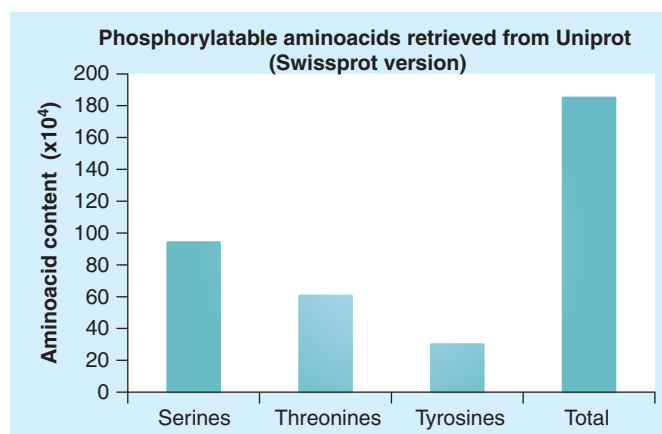
The number of identified phosphorylation sites in single experiments has increased drastically over the past decade. This development is a result of the enhanced sensitivity of mass spectrometers and increases in analysis speed, as well as improvements in sample preparation and data analysis strategies (FIGURE 2). In a seminal study in 2006, Olsen *et al.* set the bar for the identification and quantitation of great numbers of phosphorylation sites in MS-based experiments [5].

Since then a number of groups have been able to consistently report identification of tens of thousands of phosphorylation sites from both cell lines and tissues [13,21–23,35,37,51,52,74,135]. However, those impressive numbers are commonly resulting from the use of very high quantities of starting material, extensive fractionation to reduce sample complexity and subsequent increase in use of instrument time.

For all phosphoproteomics studies, a compromise has to be reached between phosphoproteome coverage, reproducibility and MS analysis time, as well as the number of replicates and conditions to analyze. Depending on the study design trade-offs have to be made between those different factors. This is not trivial, as, for example, tissue studies can be limited in sample availability. For those studies high reproducibility is of exquisite importance, to gain robust results despite hampering factors such as innate tissue and patient heterogeneity and stochastic nature of data-independent MS-based analysis. At the same time, biologically relevant phosphorylation sites, which could serve as biomarkers can be of low abundance and would be only covered in studies, reaching a certain depth.

Cell line phosphoproteomics is not as limited in sample amounts as tissue phosphoproteomics, and deep coverage has been reached in a number of recent studies. After two individual groups indicated being close to accomplishing a ‘draft’ of the complete human proteome in the last year [138,139], the question of the reachability of a ‘complete phosphoproteome’, has been revisited. We analyzed the content of serines, threonines and tyrosines, present in 20,206 reviewed protein entries from the Swiss-Prot version of Uniprot (2015–06–19), reaching a total number of 1.85 million potentially phosphorylatable amino acids in the human proteome (FIGURE 4). This number does not account for the presence of splice variants, not all amino acids being sterically accessible for phosphorylation, and phosphorylation of other amino acids than S/T/Y. Assuming around 10,000–12,000 proteins being expressed in a cell at any given time, our estimation of the total amount of possible cellular phosphorylation events is close to that of Ubersax and Ferrel. In their review, from 2007 they postulated that there are 700,000 potential phosphorylation sites in a human cell [85].

Given this impressive number, there seems to be ample opportunity for discovery of novel phosphorylation sites. This is reflected in the continuous identification of previously undescribed phosphorylation sites in deep coverage studies [21,23]. However, in the foreseeable future, it is highly unlikely that we can capture the complete human phosphoproteome or even the



**Figure 4. Phosphorylatable amino acids in the human proteome.** Serine, Threonine and Tyrosine content, as well as their sum, extracted from the SwissProt version of Uniprot (2015–06–19).

complete phosphoproteome of a specific cell at any given time. This is due not only to technical issues but also to the instant nature of phosphorylation. Phosphorylation events are highly dynamic and strongly dependent upon biological context and many exist at very low stoichiometry. Phosphorylation events – specifically tyrosine phosphorylation – can moreover occur much faster than regulation of protein levels or other PTMs. Consequently, even time-resolved phosphoproteomics analyses only capture a fraction of all possible phosphorylation sites as their abundances is highly dependent on cell-type, cell cycle status and stimuli.

In any case, we are very far from reaching the goal of a complete phosphoproteome, as most phosphoproteomics experiments rely on a relatively small number of well-established laboratory cell lines [22,23,74,77]. Moreover, tissue phosphoproteomics analyses, which include a stimulus, are still sparse and mostly constitute snapshots of a steady-state level [64,65]. While those data can serve as important compendia cataloging phosphorylation sites, putting those into a meaningful biological context will be extremely relevant. Thus, the expansion of phosphoproteome analysis to different cell types and biological cues will aid the description of a comprehensive phosphoproteome.

### Five-year view

Phosphoproteomics is now a mature analytical technology that is ready to take the next step, which is to be implemented as a standard method in cell biology laboratories, extended to tissue analyses and eventually enter the clinics. The developments of new strategies and the constant refinement of currently used laboratory protocols are making phosphoproteomics a powerful technique with great discovery power.

For the coming years, we foresee complementation of tryptic peptide-centric phosphoproteomics data with targeted top-down strategies, as well as complementary protocols using alternative proteases with different substrate cleavage specificity,

which has recently been shown to extend the accessible phosphoproteome space [31]. Those will facilitate increased phosphoproteome coverage and provide access to new levels of information such as coregulation of phosphorylation events on protein-scale.

To implement top-down strategies as sophisticated methods for medium to high-throughput phosphoproteomics different problems, relating mainly to chromatographic separation power and dynamic range have to be addressed. However, a few recent studies indicate that the field is on a good track to tackle those challenges in the coming years [24].

Despite its great promise, many issues still persist for applying phosphoproteomics in the clinics, for example, for biomarker discovery. Those are mainly related to tissue sample collection, such as ischemic artifacts [59,60], general interpatient and inter-tissue heterogeneity, and the stochastic nature of data-dependent acquisition in MS-based proteomics, as well as abundance bias of phosphoproteomics techniques. For the coming years, an introduction of standardized guidelines for sample collection and preparation will be necessary to avoid the identification of artificial phosphorylation changes. However, even with optimized protocols, the trade-off between coverage and reproducibility and the generally low amounts of sample available will likely make the standard implementation of phosphoproteomics in the clinics a great challenge for the field, which also calls for new generations of mass spectrometers with significantly improved sensitivity and sequencing speed. We foresee that advances in the field of targeted MS for phosphoproteomics, as have been recently reported [68,140], will be crucial in bridging the gap between discovery phosphoproteomics and 'single-protein' analyses and be highly useful for finding phosphorylation-based biomarkers.

Although the sequencing speed and sensitivity of mass spectrometers have increased tremendously over the last years, the boundaries of these important instrument parameters are likely improved even further in the next 5 years. Increases in sensitivity and dynamic range of instruments, in combination with

sophisticated methods to reduce sample complexity, such as powerful fragmentation strategies, will allow analysis of phosphopeptides without prior enrichment steps. This will be a crucial development to study stoichiometry of phosphorylation events in samples, which are not biased by enrichment procedures. Similarly, ion mobility-based techniques are now set to make a significant impact on phosphoproteome coverage and depth by separating multiply charged phosphopeptide species of low abundance from interfering peptides and singly charged chemical background noise.

Another important direction for the field is the development of sophisticated computational strategies to model phosphorylation-based phenotypes on a large-scale directly from quantitative phosphoproteomics datasets. Such developments are imperative for finding regulatory phosphorylation sites that are biologically meaningful within their given cellular context as opposed to simply cataloging phosphorylation events. With these developments phosphoproteomics is set to revolutionize biomarker discovery and make its way into a routine technology applied in the clinics.

#### Acknowledgements

*The authors thank Dorte Bekker-Jensen for help with figures and Lars J Jensen for his input on the manuscript.*

#### Financial & competing interests disclosure

*L von Stechow was supported by the FSS postdoctoral grant from The National Danish Research Council. L von Stechow and JV Olsen were supported in part by the research career program Sapere Aude from The National Danish Research Council. C Francavilla was supported by a long-term EMBO fellowship, Work at The Novo Nordisk Foundation Center for Protein Research (CPR) is funded in part by a generous donation from the Novo Nordisk Foundation (Grant number NNF14CC0001). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.*

#### Key issues

- MS-based phosphoproteomics is a powerful technology to study phosphorylation-dependent signaling in tissues and cells.
- Advances in MS instrumentation, especially the evolution of hybrid orbitrap instruments, have greatly improved speed and sensitivity of phosphoproteomics studies.
- Phosphoproteomics is generally performed as a bottom-up approach, typically involving tryptic digestion.
- Lower abundance of phosphopeptides, compared with their nonphosphorylated counterparts necessitates strategies to enrich phosphopeptides and/or reduce sample complexity.
- Quantitation of phosphorylation events and assessment of phosphorylation stoichiometry are crucial to determine biological relevance of identified phosphorylation sites.
- Tissue phosphoproteomics is the important next step for the clinical application of phosphoproteomics, but is challenged by issues relating to sample collection such as post-mortem events and at this point mainly represents snapshots of steady-state conditions.
- Studying crosstalk between phosphorylation and other posttranslational modifications (PTMs) is highly relevant, but functionally limited by the commonly used bottom-up approach.
- Bioinformatics tools are necessary to identify the biological context in which phosphorylation events function.

## References

Papers of special note have been highlighted as:

• of interest

•• of considerable interest

1. Olsen JV, Mann M. Status of Large-scale Analysis of Post-translational Modifications by Mass Spectrometry. *Mol Cell Proteom* 2013;12(12):3444-52
2. Solari FA, Dell'Aica M, Sickmann A, Zahedi RP. Why phosphoproteomics is still a challenge. *Mol Biosyst* 2015;11(6):1487-93
3. Beausoleil SA, Villen J, Gerber SA, et al. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* 2006;24(10):1285-92
4. Olsen JV, Mann M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci USA* 2004;101(37):13417-22
5. Olsen JV, Blagoev B, Gnäd F, et al. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 2006;127(3):635-48
- **This paper shows the first large-scale application of quantitative phosphoproteomics to dissect a dynamic signaling pathway and made significant impact on the field of phosphoproteomics.**
6. Syka JE, Coon JJ, Schroeder MJ, et al. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA* 2004;101(26):9528-33
7. Olsen JV, Macek B, Lange O, et al. Higher-energy C-trap dissociation for peptide modification analysis. *Nat Methods* 2007;4(9):709-12
8. Nagaraj N, D'Souza RC, Cox J, et al. Feasibility of large-scale phosphoproteomics with higher energy collisional dissociation fragmentation. *J Proteome Res* 2010;9(12):6786-94
9. Engholm-Keller K, Larsen MR. Technologies and challenges in large-scale phosphoproteomics. *Proteomics* 2013;13(6):910-31
10. Jedrychowski MP, Huttlin EL, Haas W, et al. Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol Cell Proteomics* 2011;10(12):M111 009910
11. Zhou H, Low TY, Hennrich ML, et al. Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol Cell Proteomics* 2011;10(10):M110 006452
12. Marx H, Lemeer S, Schliep JE, et al. A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nat Biotechnol* 2013;31(6):557-64
13. de Graaf EL, Giansanti P, Altelaar AF, Heck AJ. Single-step enrichment by Ti4+-IMAC and label-free quantitation enables in-depth monitoring of phosphorylation dynamics with high reproducibility and temporal resolution. *Mol Cell Proteomics* 2014;13(9):2426-34
- **This study indicates use of a Ti4+enrichment and label-free quantitation and might present a valuable strategy for high-throughput analyses of cell lines and tissues.**
14. Bridon G, Bonneil E, Muratore-Schroeder T, et al. Improvement of phosphoproteome analyses using FAIMS and decision tree fragmentation. application to the insulin signaling pathway in *Drosophila melanogaster* S2 cells. *J Proteome Res* 2012;11(2):927-40
15. Brunner AM, Lossel P, Liu F, et al. Benchmarking multiple fragmentation methods on an orbitrap fusion for top-down phospho-proteome characterization. *Anal Chem* 2015;87(8):4152-8
16. Frese CK, Zhou H, Taus T, et al. Unambiguous phosphosite localization using electron-transfer/higher-energy collision dissociation (ETHCD). *J Proteome Res* 2013;12(3):1520-5
17. Venne K, Bonneil E, Eng K, Thibault P. Improvement in peptide detection for proteomics analyses using nanoLC-MS and high-field asymmetry waveform ion mobility mass spectrometry. *Anal Chem* 2005;77(7):2176-86
18. Saba J, Bonneil E, Pomies C, et al. Enhanced Sensitivity in Proteomics Experiments Using FAIMS Coupled with a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *J Proteome Res* 2009;8(7):3355-66
19. Helm D, Vissers JP, Hughes CJ, et al. Ion mobility tandem mass spectrometry enhances performance of bottom-up proteomics. *Mol Cell Proteomics* 2014;13(12):3709-15
20. Creese AJ, Smart J, Cooper HJ. Large-scale analysis of peptide sequence variants: The case for high field asymmetric waveform ion mobility spectrometry. *Anal Chem* 2013;85:4836-43
21. Huttlin EL, Jedrychowski MP, Elias JE, et al. A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* 2010;143(7):1174-89
- **This study constitutes the first large-scale analyses of phosphorylation in tissues.**
22. Olsen JV, Vermeulen M, Santamaria A, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci signal* 2010;3(104):ra3
- **This study shows for the first time, sophisticated stoichiometry determination on a proteome-wide scale without introducing further experimental steps.**
23. Sharma K, D'Souza RC, Tyanova S, et al. Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell reports* 2014;8(5):1583-94
- **This study reached so far unprecedented depth of phosphorylation coverage.**
24. Tran JC, Zamdborg L, Ahlf DR, et al. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* 2011;480(7376):254-8
- **This study is an exceptional example of a global Top-Down analysis of proteins and their posttranslational modifications.**
25. Mann M, Hendrickson RC, Pandey A. Analysis of proteins and proteomes by mass spectrometry. *Annu Rev Biochem* 2001;70:437-73
26. Switzer L, Giera M, Niessen WM. Protein digestion: an overview of the available techniques and recent developments. *J Proteome Res* 2013;12(3):1067-77
27. Olsen JV, Ong SE, Mann M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol Cell Proteomics* 2004;3(6):608-14
28. Gilmore JM, Kettenbach AN, Gerber SA. Increasing phosphoproteomic coverage through sequential digestion by complementary proteases. *Anal Bioanal Chem* 2012;402(2):711-20
29. Dickhut C, Feldmann I, Lambert J, Zahedi RP. Impact of digestion conditions on phosphoproteomics. *J Proteome Res* 2014;13(6):2761-70
30. Hamady M, Cheung TH, Tufo H, Knight R. Does protein structure influence trypsin miscleavage? Using structural properties to predict the behavior of related proteins. *IEEE Eng Med Biol Mag* 2005;24(3):58-66

31. Giansanti P, Aye TT, van den Toorn H, et al. An Augmented Multiple-Protease-Based Human Phosphopeptide Atlas. *Cell reports* 2015; 11(11):1834-43
32. Beltran L, Cutillas PR. Advances in phosphopeptide enrichment techniques for phosphoproteomics. *Amino Acids* 2012; 43(3):1009-24
33. Hunter T, Sefton BM. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci USA* 1980;77(3):1311-15
34. Boersema PJ, Foong LY, Ding VM, et al. In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. *Mol Cell Proteomics* 2010;9(1):84-99
35. Engholm-Keller K, Birck P, Storling J, et al. TiSH—a robust and sensitive global phosphoproteomics strategy employing a combination of TiO<sub>2</sub>, SIMAC, and HILIC. *J Proteomics* 2012;75(18):5749-61
36. Thingholm TE, Jensen ON, Robinson PJ, Larsen MR. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol Cell Proteomics* 2008;7(4):661-71
37. Batth TS, Francavilla C, Olsen JV. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J Proteome Res* 2014;13(12):6176-86
38. Engholm-Keller K, Hansen TA, Palmisano G, Larsen MR. Multidimensional strategy for sensitive phosphoproteomics incorporating protein prefractionation combined with SIMAC, HILIC, and TiO<sub>2</sub> (2) chromatography applied to proximal EGF signaling. *J Proteome Res* 2011; 10(12):5383-97
39. Yu Z, Han G, Sun S, et al. Preparation of monodisperse immobilized Ti(4+) affinity chromatography microspheres for specific enrichment of phosphopeptides. *Anal Chim Acta* 2009;636(1):34-41
40. Zhou H, Ye M, Dong J, et al. Specific phosphopeptide enrichment with immobilized titanium ion affinity chromatography adsorbent for phosphoproteome analysis. *J Proteome Res* 2008;7(9):3957-67
41. Zhou H, Ye M, Dong J, et al. Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography. *Nat Protoc* 2013;8(3): 461-80
42. Ruprecht B, Koch H, Medard G, et al. Comprehensive and reproducible phosphopeptide enrichment using iron immobilized metal ion affinity chromatography (Fe-IMAC) columns. *Mol Cell Proteomics* 2015;14(1):205-15
43. Iliuk AB, Martin VA, Alicie BM, et al. In-depth analyses of kinase-dependent tyrosine phosphoproteomes based on metal ion-functionalized soluble nanoparticles. *Mol Cell Proteomics* 2010;9(10):2162-72
44. Searleman AC, Iliuk AB, Collier TS, et al. Tissue phosphoproteomics with PolyMAC identifies potential therapeutic targets in a transgenic mouse model of HER2 positive breast cancer. *Electrophoresis* 2014;35(24): 3463-9
45. Kinoshita E, Kinoshita-Kikuta E, Koike T. Advances in Phos-tag-based methodologies for separation and detection of the phosphoproteome. *Biochim Biophys Acta* 2015;1854(6):601-8
46. Rainer M, Bonn GK. Enrichment of phosphorylated peptides and proteins by selective precipitation methods. *Bioanalysis* 2015;7(2):243-52
47. Matsuoka S, Ballif BA, Smogorzewska A, et al. ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* 2007;316(5828): 1160-6
48. Machida K, Khenkhar M, Nollau P. Deciphering phosphotyrosine-dependent signaling networks in cancer by SH2 profiling. *Genes Cancer* 2012;3(5-6): 353-61
49. Galan JA, Geraghty KM, Lavoie G, et al. Phosphoproteomic analysis identifies the tumor suppressor PDCD4 as a RSK substrate negatively regulated by 14-3-3. *Proc Natl Acad Sci USA* 2014;111(29): E2918-27
50. Loroeh S, Zahedi RP, Sickmann A. Highly sensitive phosphoproteomics by tailoring solid-phase extraction to electrostatic repulsion-hydrophilic interaction chromatography. *Anal Chem* 2015;87(3): 1596-604
51. Lundby A, Secher A, Lage K, et al. Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nat Commun* 2012;3:876
52. Rigbolt KT, Prokhorova TA, Akimov V, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal* 2011;4(164):rs3
53. Kettenbach AN, Gerber SA. Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. *Anal Chem* 2011;83(20): 7635-44
54. Masuda T, Sugiyama N, Tomita M, Ishihama Y. Microscale phosphoproteome analysis of 10,000 cells from human cancer cell lines. *Anal Chem* 2011;83(20): 7698-703
- **This study highlights strategies for analysis and quantitation of phosphorylation sites from minute amounts of input material.**
55. Lundby A, Lage K, Weinert BT, et al. Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. *Cell reports* 2012;2(2): 419-31
56. Liu NQ, Braakman RB, Stingl C, et al. Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue. *J Mammary Gland Biol Neoplasia* 2012;17(2):155-64
57. Casado P, Bilanges B, Rajeev V, et al. Environmental stress affects the activity of metabolic and growth factor signaling networks and induces autophagy markers in MCF7 breast cancer cells. *Mol Cell Proteomics* 2014;13(3):836-48
58. Kanshin E, Tyers M, Thibault P. Sample collection method bias effects in quantitative phosphoproteomics. *J Proteome Res* 2015; 14(7):2998-3004
59. Mertins P, Yang F, Liu T, et al. Ischemia in tumors induces early and sustained phosphorylation changes in stress kinase pathways but does not affect global protein levels. *Mol Cell Proteomics* 2014;13(7): 1690-704
60. Gajadhar AS, Johnson H, Slebos RJ, et al. Phosphotyrosine signaling analysis in human tumors is confounded by systemic ischemia-driven artifacts and intra-specimen heterogeneity. *Cancer Res* 2015;75(7): 1495-503
61. Boren M. Sample preservation through heat stabilization of proteins: principles and examples. *Methods Mol Biol* 2015;1295: 21-32
62. Wakabayashi M, Yoshihara H, Masuda T, et al. Phosphoproteome analysis of formalin-fixed and paraffin-embedded tissue sections mounted on microscope slides. *J Proteome Res* 2014;13(2):915-24



63. Zhai B, Villen J, Beausoleil SA, et al. Phosphoproteome analysis of *Drosophila melanogaster* embryos. *J Proteome Res* 2008;7(4):1675-82
64. Lundby A, Andersen MN, Steffensen AB, et al. In vivo phosphoproteomics analysis reveals the cardiac targets of beta-adrenergic receptor signaling. *Sci signal* 2013;6(278):rs11
65. Monetti M, Nagaraj N, Sharma K, Mann M. Large-scale phosphosite quantification in tissues by a spike-in SILAC method. *Nat Methods* 2011;8(8):655-8
66. Weigand S, Herting F, Maisel D, et al. Global quantitative phosphoproteome analysis of human tumor xenografts treated with a CD44 antagonist. *Cancer Res* 2012;72(17):4329-39
67. Corradini E, Vallur R, Raaijmakers LM, et al. Alterations in the cerebellar (Phospho) proteome of a cyclic guanosine monophosphate (cGMP)-dependent protein kinase knockout mouse. *Mol Cell Proteomics* 2014;13(8):2004-16
68. Narumi R, Murakami T, Kuga T, et al. A strategy for large-scale phosphoproteomics and SRM-based validation of human breast cancer tissue samples. *J Proteome Res* 2012;11(11):5311-22
- **This study highlights the power of phosphoproteomics as a tool for developing hypothesis about biological phenotypes and their validation *in vitro* and *in vivo*.**
69. Dephoure N, Gygi SP. Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Sci.Signal* 2012;5(217):rs2
70. Everley RA, Kunz RC, McAllister FE, Gygi SP. Increasing throughput in targeted proteomics assays: 54-plex quantitation in a single mass spectrometry run. *Anal.Chem* 2013;85(11):5340-6
71. Francavilla C, Rigbolt KT, Emdal KB, et al. Functional proteomics defines the molecular switch underlying FGF receptor trafficking and cellular outputs. *Mol Cell* 2013;51(6):707-22
72. Pan C, Olsen JV, Daub H, Mann M. Global effects of kinase inhibitors on signaling networks revealed by quantitative phosphoproteomics. *Mol Cell Proteomics* 2009;8(12):2796-808
73. Bennetzen MV, Larsen DH, Bunkenborg J, et al. Site-specific phosphorylation dynamics of the nuclear proteome during the DNA damage response. *Mol Cell Proteomics* 2010;9(6):1314-23
74. Beli P, Lukashchuk N, Wagner SA, et al. Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol Cell* 2012;46(2):212-25
75. Bensimon A, Schmidt A, Ziv Y, et al. ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage. *Sci signal* 2010;3(151):rs3
76. Pines A, Kelstrup CD, Vrouwe MG, et al. Global Phosphoproteome Profiling Reveals Unanticipated Networks Responsive to Cisplatin Treatment of Embryonic Stem Cells. *Mol Cell Biol* 2011;31(24):4964-77
77. Mertins P, Udeshi ND, Clauser KR, et al. iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Mol Cell Proteomics* 2012;11(6):M111-014423
78. Cox J, Hein MY, Luber CA, et al. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* 2014;13(9):2513-26
79. Wu R, Haas W, Dephoure N, et al. A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nat Methods* 2011;8(8):677-83
80. Tsai CF, Wang YT, Yen HY, et al. Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nat comm* 2015;6:6622
81. Gerber SA, Rush J, Stemman O, et al. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc Natl Acad Sci USA* 2003;100(12):6940-5
82. Curran TG, Zhang Y, Ma DJ, et al. MARQUIS: a multiplex method for absolute quantification of peptides and posttranslational modifications. *Nat comm* 2015;6:5924
83. Steen H, Mann M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 2004;5(9):699-711
84. Beltrao P, Bork P, Krogan NJ, van Noort V. Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* 2013;9:714
85. Ubersax JA, Ferrell JE Jr. Mechanisms of specificity in protein phosphorylation. *Nat Rev Mol Cell Biol* 2007;8(7):530-41
86. Zheng Y, Zhang C, Croucher DR, et al. Temporal regulation of EGF signalling networks by the scaffold protein Shc1. *Nature* 2013;499(7457):166-71
87. Yu Y, Yoon SO, Poulogiannis G, et al. Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science* 2011;332(6035):1322-6
88. Hsu PP, Kang SA, Rameseder J, et al. The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* 2011;332(6035):1317-22
89. Emdal KB, Pedersen AK, Bekker-Jensen DB, et al. Temporal proteomics of NGF-TrkA signaling identifies an inhibitory role for the E3 ligase Cbl-b in neuroblastoma cell differentiation. *Sci. Signal* 2015;8(374):ra40
90. Terfve C, Saez-Rodriguez J. Modeling signaling networks using high-throughput phospho-proteomics. *Adv Exp Med Biol* 2012;736:19-57
91. Kanshin E, Bergeron-Sandoval LP, Isik SS, et al. A cell-signaling network temporally resolves specific versus promiscuous phosphorylation. *Cell reports* 2015;10(7):1202-14
92. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet* 2000;25(1):25-9
93. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27-30
94. Huang DW, Sherman BT, Tan Q, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res* 2007;35:W169-75
95. Breuer K, Foroushani AK, Laird MR, et al. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 2013;41:D1228-33
96. Franceschini A, Szklarczyk D, Frankild S, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 2013;41:D808-15
97. Kerrien S, Aranda B, Breuza L, et al. The IntAct molecular interaction database in 2012. *Nucleic Acids Res* 2012;40:D841-6
98. Licata L, Briganti L, Peluso D, et al. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012;40: D857-61
99. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for

- integrated models of biomolecular interaction networks. *Genome Res* 2003; 13(11):2498-504
100. Klammer M, Godl K, Tebbe A, Schaab C. Identifying differentially regulated subnetworks from phosphoproteomic data. *BMC Bioinformatics* 2010;11:351
101. Casado P, Rodriguez-Prados JC, Cosulich SC, et al. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci signal* 2013; 6(268):rs6
102. Vaga S, Bernardo-Faura M, Cokelaer T, et al. Phosphoproteomic analyses reveal novel cross-modulation mechanisms between two signaling pathways in yeast. *Mol Syst Biol* 2014;10:767
103. Landry CR, Levy ED, Michnick SW. Weak functional constraints on phosphoproteomes. *Trends Genet* 2009; 25(5):193-7
104. Beltrao P, Trinidad JC, Fiedler D, et al. Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. *PLoS Biol* 2009;7(6):e1000134
105. Beltrao P, Albanese V, Kenner LR, et al. Systematic functional prioritization of protein posttranslational modifications. *Cell* 2012;150(2):413-25
- **This study describes an interesting bioinformatics strategy to functionally prioritize phosphorylation sites from a compendium of high-throughput studies.**
106. Hunter T. The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol Cell* 2007;28(5):730-8
107. Swaney DL, Beltrao P, Starita L, et al. Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nat Methods* 2013;10(7): 676-82
108. Tan CS, Bader GD. Phosphorylation sites of higher stoichiometry are more conserved. *Nat Methods* 2012;9(4):317
109. Weinert BT, Wagner SA, Horn H, et al. Proteome-wide mapping of the *Drosophila* acetylome demonstrates a high degree of conservation of lysine acetylation. *Sci signal* 2011;4(183):ra48
110. Weinert BT, Iesmantavicius V, Moustafa T, et al. Acetylation dynamics and stoichiometry in *Saccharomyces cerevisiae*. *Mol Syst Biol* 2014;10:716
111. Tan CS, Bodenmiller B, Pasculescu A, et al. Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. *Sci signal* 2009;2(81):ra39
112. Wang T, Kettenbach AN, Gerber SA, Bailey-Kellogg C. MMFP: a maximal motif finder for phosphoproteomics datasets. *Bioinformatics* 2012;28(12): 1562-70
113. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell* 2013;153(1):17-37
114. Cohen P. Protein kinases—the major drug targets of the twenty-first century? *Nat Rev Drug Discov* 2002;1(4):309-15
115. Cohen P, Alessi DR. Kinase drug discovery—what's next in the field? *ACS Chem Biol* 2013;8(1):96-104
116. Dinkel H, Chica C, Via A, et al. Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res* 2011;36:D261-7
117. Linding R, Jensen LJ, Pasculescu A, et al. NetworkKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res* 2008;36:D695-9
118. Schwartz D, Gygi SP. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat Biotechnol* 2005;23(11):1391-8
- **This study is an important example of comprehensive kinome and phosphatome profiling to identify kinase-substrate relationships.**
119. Colaert N, Helsens K, Martens L, et al. Improved visualization of protein consensus sequences by iceLogo. *Nat Methods* 2009; 6(11):786-7
120. Wagner MJ, Stacey MM, Liu BA, Pawson T. Molecular mechanisms of SH2- and PTB-domain-containing proteins in receptor tyrosine kinase signaling. *Cold Spring Harb Perspect Biol* 2013;5(12): a008987
121. Bodenmiller B, Wanka S, Kraft C, et al. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci signal* 2010;3(153):rs4
122. Tagliabracci VS, Wiley SE, Guo X, et al. A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* 2015; 161(7):1619-32
123. Holt LJ, Tuch BB, Villen J, et al. Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* 2009;325(5948):1682-6
- **This study shows an interesting approach for identification of kinase-substrate relationships.**
124. Lourido S, Jeschke GR, Turk BE, Sibley LD. Exploiting the unique ATP-binding pocket of toxoplasma calcium-dependent protein kinase 1 to identify its substrates. *ACS Chem Biol* 2013;8(6):1155-62
125. Becher I, Savitski MM, Savitski MF, et al. Affinity profiling of the cellular kinome for the nucleotide cofactors ATP, ADP, and GTP. *ACS Chem Biol* 2013;8(3):599-607
126. van de Waterbeemd M, Lossl P, Gautier V, et al. Simultaneous assessment of kinetic, site-specific, and structural aspects of enzymatic protein phosphorylation. *Angew Chem Int Ed Engl* 2014;53(36):9660-4
127. Kubota K, Anjum R, Yu Y, et al. Sensitive multiplexed analysis of kinase activities and activity-based kinase identification. *Nat Biotechnol* 2009;27(10):933-40
128. Yu Y, Anjum R, Kubota K, et al. A site-specific, multiplexed kinase activity assay using stable-isotope dilution and high-resolution mass spectrometry. *Proc Natl Acad Sci USA* 2009;106(28):11606-11
129. Xue L, Geahlen RL, Tao WA. Identification of direct tyrosine kinase substrates based on protein kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* 2013;12(10):2969-80
130. Xue L, Wang P, Cao P, et al. Identification of extracellular signal-regulated kinase 1 (ERK1) direct substrates using stable isotope labeled kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* 2014;13(11):3199-210
131. Kettenbach AN, Wang T, Faherty BK, et al. Rapid determination of multiple linear kinase substrate motifs by mass spectrometry. *Chem Biol* 2012;19(5): 608-18
132. Soufi B, Soares NC, Ravikumar V, Macek B. Proteomics reveals evidence of cross-talk between protein modifications in bacteria: focus on acetylation and phosphorylation. *Curr Opin Microbiol* 2012;15(3):357-63
133. Ruthenburg AJ, Li H, Patel DJ, Allis CD. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* 2007;8(12):983-94
134. Brooks CL, Gu W. Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Curr Opin Cell Biol* 2003;15(2):164-71
135. Mertins P, Qiao JW, Patel J, et al. Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat Methods* 2013;10(7):634-7

136. Iesmantavicius V, Weinert BT, Choudhary C. Convergence of ubiquitylation and phosphorylation signaling in rapamycin-treated yeast cells. *Mol Cell Proteomics* 2014;13(8):1979-92
137. Van Roey K, Dinkel H, Weatheritt RJ, et al. The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci Signal* 2013;6(269):rs7
138. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. *Nature* 2014;509(7502):575-81
139. Wilhelm M, Schlegl J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. *Nature* 2014;509(7502):582-7
140. Parker BL, Yang G, Humphrey SJ, et al. Targeted phosphoproteomics of insulin signaling using data-independent acquisition mass spectrometry. *Sci Signal* 2015;8(380):rs6